A NOVEL MULTI-STAGE MOTION VECTOR PROCESSING METHOD FOR MOTION COMPENSATED FRAME INTERPOLATION

Ai-Mei Huang and Truong Nguyen

Video Processing Lab
ECE Dept, UCSD, La Jolla, CA 92093
E-mail: aihuang@ucsd.edu, nguyent@ece.ucsd.edu

ABSTRACT

In this paper, a novel multi-stage motion vector processing algorithm at the decoder is proposed for motion compensated frame interpolation. We address the problems of discontinuous edges and deformed structures in an interpolated frame by explicitly considering reliability of each received motion vector. By hierarchically refining motion vectors with different block sizes, the proposed method is capable of preserving structure information. Experimental results show that the proposed scheme outperforms other methods in terms of visual quality and PSNR, and it is also robust when video sequences have complex scenes and fast motion.

Index Terms— motion compensated frame interpolation, frame rate up conversion, motion vector processing, residual energy

1. INTRODUCTION

Motion-compensated frame interpolation (MCFI) has recently been studied to improve temporal quality by increasing the frame rate at the decoder. MCFI interpolates the skipped frames by averaging forward and backward motion compensated predictions based on the assumption that objects move along the motion trajectory. The received motion vectors (MVs) are often divided by two and directly used for frame interpolation, which is also called the *direct* MCFI. However, the received MVs are often unreliable and do not represent the actual motion. This is because the received motion vector field (MVF) is usually generated using block-based motion estimation at the encoder by minimizing prediction errors, rather than finding true motion. As a result, MCFI that directly uses the received MVs often suffers from annoying artifacts such as blockiness and ghost effect.

To solve this problem, a number of works have been proposed to obtain a better MVF for MCFI. Shinya and Akira proposed using a larger block size for global motion regions while a smaller block size is used for local motion regions [1]. The works in [2] and [3] presented MV processing techniques to simply remove MV outliers and/or refine MVs from its neighborhood. However, those MV processing methods that remove outliers using vector median filter (VMF) or refine MVs using smaller block sizes can only perform well when the video has smooth and regular motion. They usually do not consider the edge continuity of the objects and often fail on the motion boundaries. In addition, intra-coded MBs make frame interpolation more difficult as their MVs are not available and need to be estimated. Therefore, frame interpolation for compressed video still remains a challenging problem as the artifacts due to the use of improper MVs are often generated.

In our previous work, we proposed two different MV processing methods, which refines MVs using a constrained VMF [4] or selects a single MV for a merged MB group from the neighborhood [5]. In this paper, we combine these two methods and further propose a novel multi-stage yet low-complexity MV processing method to not only preserve the object structure information but also produce a smoother MVF. As suggested in [4] and [5], we first identify unreliable MVs using the residual energy information. Then, before refining those unreliable MVs based on smaller blocks, we merge MBs that have unreliable MVs by analyzing the distribution of the residual energies. After assigning a single MV to each merged group, we further use an effective MV refinement similar to [4] that adaptively adjusts unreliable MVs in a smaller block size and use MV smoothing as in [3] to produce an even finer MVF. In addition, adaptively selecting forward and backward predictions based on the motion and using chrominance information for MV reliability classification and MV processing are also addressed. Color information is found very useful to identify and correct unreliable MVs and has not explicitly been considered in the literature. The simulation results show that the proposed multi-stage MV processing method can significantly improve visual quality and outperform the conventional MV processing methods.

The rest of this paper is organized as follows. We first illustrate how to use the correlation between MV reliability and the energy of residual signals to create a MV reliability map and a MB merging map in Section 2. The proposed MV processing method is described in details in Section 3. The simulation results on various video sequences are demonstrated in Section 4. Finally, some conclusions are given in Section 5.

2. PREDICTION RESIDUAL ENERGY ANALYSIS

In [4], we have discussed that there exists a strong correlation between MV reliability and its associated residual energy. Artifacts such as blockiness and deformed structures appear at the areas where the residual energies are high or the areas where no MV is available, as shown in Figs. 1(a) and (b). In Fig. 1(a), the structure of the white lines cannot be maintained. It is because the MVs are usually estimated separately such that those blocks may not have the same MV to be perfectly assembled together. However, if we look closely at how these high residual energies are distributed, we can roughly tell where object edges are located. That is, we should assign a single MV for those MBs that are connected by high residual energies. By doing so, the structure can first be maintained before any further MV refinement. Therefore, by using residual energy information, we can create a MV reliability map and a MB merging map to facilitate MV processing. The MV reliability map denotes where unreliable MVs

---

This work is supported by Conexant Inc. and matching fund from UC Discovery program.

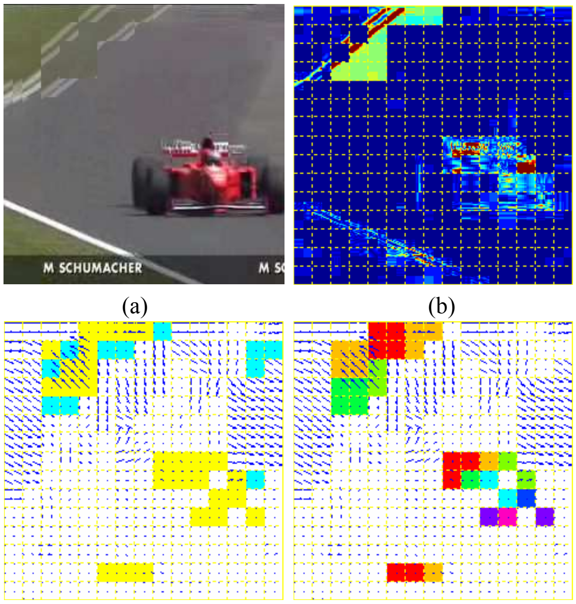
(a) (b)
(c) (d)

Fig. 1. (a) Interpolation result of the frame 56 of FORMULA 1 sequence using direct MCFI from reconstructed frames 55 and 57. (b) Residual energy of the reconstructed frame 57. (c) MV reliability classification map. Unreliable and reliable MVs are marked in yellow and white colors, respectively. Intra-coded MBs are marked in cyan color. (d) MB merging map.

are and the MB merging map tells whether the neighboring MBs should be grouped together to maintain the integrity of the entire moving object.

The reliability levels of received MVs are determined as follows. Let $\mathbf{v}_{m,n}$ denote the MV of each 8×8 block. We classify $\mathbf{v}_{m,n}$ into three different reliability levels, *reliable*, *possibly reliable*, and *unreliable*, based on its residual energy, the reliability level of its neighboring blocks and the coding type. For a MB with only one MV, we simply assign the same MV to all four 8×8 blocks. The residual energy for each block $\mathbf{b}_{m,n}$ is the sum of the absolute value of each received prediction error of each pixel. Since MVs estimated in luminance domain may result in mismatched color, we include chrominance information in residual energy calculation to identify those incorrect MVs. $E_{m,n}$ can then be represented as follows:

$$E_{m,n} = \sum_{(i,j)\in\mathbf{b}^{Y}_{m,n}} |r_Y(i,j)| + \alpha\cdot\Big(\sum_{(i,j)\in\mathbf{b}^{Cb}_{m,n}} |r_{Cb}(i,j)| + \sum_{(i,j)\in\mathbf{b}^{Cr}_{m,n}} |r_{Cr}(i,j)|\Big) \quad (1)$$

where $r_Y(i,j)$, $r_{Cb}(i,j)$, and $r_{Cr}(i,j)$ are the reconstructed residual signals of Y, Cb and Cr components of the block, $\mathbf{b}_{m,n}$, respectively. α is the weight used to emphasize the degree of color difference. Please note that there is no additional computation of using such information other than Eqn. (1).

We then compare $E_{m,n}$ with a predefined threshold, ε_1, to determine if $\mathbf{v}_{m,n}$ is unreliable. If $E_{m,n}$ is greater than or equal to ε_1, it will be considered as *unreliable* and inserted into the reliability level set L_1. For intra-coded MBs, since they do not have MVs, we consider them as unreliable and put them in L_1. Once an unreliable

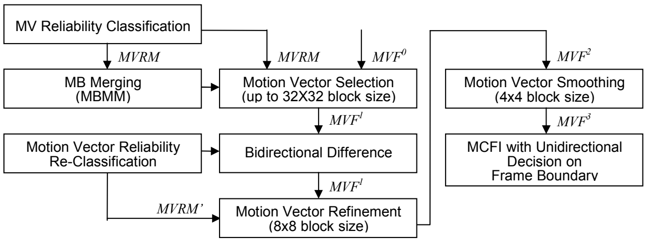

Fig. 2. Block diagram of the proposed algorithms. MVF^k denotes the updated motion vector field after each process.

MV is identified, the neighboring MVs in the same MB and eight adjacent MBs will be classified as *possibly reliable* and be placed into the second reliability level set L_2, if their residual energies are less than ε_1. The reason why we consider L_2 is that there may exist a motion boundary on those MBs. To ensure that all MVs used for frame interpolation are reliable, we mark these MVs as possibly reliable and they will be further verified in a later stage of the MV correction process. For those MVs that are not classified yet and their $E_{m,n}$ are less than ε_1, they will be classified as *reliable* and placed into the third reliability level set L_3. Therefore, we can create a MV reliability map (MVRM) by assigning the reliability level to each MV as follows:

$$MVRM(m,n) = \begin{cases} L_1, & \text{if } E_{m,n} \geq \varepsilon_1, \\ L_2, & \text{if any MV in the same MB or} \\ & \text{in the adjacent MBs} \in L_1, \\ L_3, & \text{otherwise.} \end{cases}$$

Fig. 1(c) demonstrates the MV reliability map based on Fig. 1(b). Since the luminance values of the pavement and grass are similar, some of the wrong MVs can only be detected by chrominance residues instead of luminance residues. As expected, we can successfully identify the regions where frame interpolation is most likely to fail by classifying the MV reliability.

After classifying the reliability of each MV, we analyze the *connectivity* of the unreliable MVs in MVRM and create a MB merging map. The merging process is performed on a MB basis using MVRM, and all MBs that contain unreliable MVs will be examined in a raster scan order. If a MB that has unreliable MVs connecting to other unreliable MVs in vertical, horizontal or diagonal directions in adjacent MBs, these MBs will be merged. We choose 32×32 block size as the maximum block size for merging. If adjacent MBs have been merged, this MB will remain as a single 16×16 block. If two adjacent MBs that have unreliable MVs but they are not next to each other, those two MBs will not be merged. In this merging process, intra-coded MBs are automatically included as their MVs are considered unreliable. However, the diagonal direction is not considered for intra-intra MB merging. It is because the possibility for two diagonal intra-coded MBs belonging to the same object is lower. A MB merging map (MBMM) can then be created by assigning a unique number to the MBs that are merged, indicating that they should be considered together to find a single MV in the MV processing stage. Fig. 1(d) shows the MB merging map, where all the MBs in the same merged group is marked in the same color. Red color is the default color if a merged group has no other merged groups next to it. From Fig. 1(b) and Fig. 1(d), we can see that some blocks with high residual energies have been grouped together (i.e., those along the white lines).

3. THE PROPOSED MULTI-STAGE MOTION VECTOR PROCESSING METHOD

The block diagram of the proposed method is illustrated in Fig. 2. First, we use MV selection as described in [5] to find the best MV for each merged group based on the MV reliability map, the MB merging map, and the received MVF, MVF^0. The MV candidates for MV selection are formed from the reliable MVs of the MBs in the merged group and their adjacent MBs. The best MV, $\mathbf{v}_b^*$, is chosen from these candidates by minimizing the averaged absolute bidirectional prediction difference (ABPD) between the forward and backward predictions.

$$\mathbf{v}_b^* = \arg\min_{\mathbf{v}\in S}(ABPD(\mathbf{v})), \tag{2}$$

where

$$\begin{aligned} ABPD(\mathbf{v}) \quad = \quad & \frac{1}{N_{C_u}} \sum_{i,j\in C_u} |f_{t-1}(i+\frac{1}{2}v_x, j+\frac{1}{2}v_y) - \\ & f_{t+1}(i-\frac{1}{2}v_x, j-\frac{1}{2}v_y)|. \end{aligned}$$

S denotes the set of the MV candidates and C_u denotes the merged group. Again, we consider both luminance and chrominance information in Eqn.(2) and use the same weighting factor as in Eqn.(1). Before assigning the selected MV to the merged MBs in C_u, we need to check if its ABPD is less than a threshold ε_2. If yes, all MVs in C_u will be replaced by the new MV, $\mathbf{v}_b^*$, and marked done. $\mathbf{v}_b^*$ then will be used to update the received MVF^0 to MVF^1. Otherwise, we drop the selected MV and skip this merged group temporarily to see if there will be better MVs from neighboring corrected merged groups in the next iteration. That is, we wait until a proper MV propagates to its neighborhood and the whole process stops until all merged groups have been assigned new MVs. In our simulation, we set the iteration number to be 2 and increase ε_2 to a very high value in the second iteration so that all merged blocks will certainly be assigned new MVs.

By assigning a single MV to each merged group, we can maintain edge information and the integrity of moving objects. However, since a MB in high residual area will only have one major motion, if the MB consists of multiple motion, regions having different motion can produce visual artifacts. Those unreliable MV can be easily detected by calculating bidirectional prediction difference (BPD).

As shown in Fig. 2, during MV selection, we further reclassify MV reliability based on BPD resulted from the selected MV. Besides, BPD of those possibly reliable MVs will be checked to see if they are truly reliable during reclassification. $BPD(m,n)$ of each 8×8 block is obtained by simply summing up difference error of each pixel with the same criteria described in Eqn. (1). If it is higher than a threshold ε_3, then the MV $\mathbf{v}_{m,n}^*$ will be classified as unreliable and put in L_1. Otherwise, those MVs will be classified as reliable and put in L_3.

$$MVRM(m,n) = \begin{cases} L_1, & \text{if } BPD(m,n) \geq \varepsilon_3, \\ L_3, & \text{if } BPD(m,n) < \varepsilon_3. \end{cases} \tag{3}$$

In this stage, we only consider two reliability levels for MVRM. MVs of L_2 will be classified into L_1 or L_3. By doing this, we can successfully differentiate improper motion and then refine those detected unreliable MVs based on smaller block size of 8×8 in the subsequent motion refinement stage.

Table 1. Weight values for forward and backward predictions on frame boundary.

$m=1$	w_{t-1}	w_{t+1}	$n=1$	w_{t-1}	w_{t+1}
$v_y \leq 0$	0	1	$v_y \leq 0$	0	1
$v_y > 0$	1	0	$v_y > 0$	1	0
$m=M$	w_{t-1}	w_{t+1}	$n=N$	w_{t-1}	w_{t+1}
$v_x \leq 0$	1	0	$v_x \leq 0$	1	0
$v_x > 0$	0	1	$v_x > 0$	0	1

For those unreliable MVs in the updated MVRM, we correct them using a *reliability and similarity constrained vector median filter* in the following:

$$v_{m,n}^* = \arg\min_{v\in S} \sum_{i=m-1}^{m+1} \sum_{j=n-1}^{n+1} w_{i,j}\, \|v - v_{i,j}\|, \tag{4}$$

where

$$w_{i,j} = \begin{cases} 0, \; if\ MVRM(i,j) = L_1, \\ 1, \; if\ MVRM(i,j) = L_3 \text{ and } d_{i,j} > \varepsilon_4. \end{cases}$$

S contains the neighboring MVs centered at $\mathbf{v}_{m,n}$, and $d_{i,j}$ denotes the distance between $\mathbf{v}_{i,j}$ and the centered MV, $\mathbf{v}_{m,n}$ using the angular difference.

$$d_{i,j} = 1 - \frac{\mathbf{v}_{m,n}\cdot\mathbf{v}_{i,j}}{|\mathbf{v}_{m,n}||\mathbf{v}_{i,j}|} = 1 - \cos\theta$$

where θ is the angle between $\mathbf{v}_{i,j}$ and $\mathbf{v}_{m,n}$. The distance is used for measuring the dissimilarity of the candidate MVs and the original MV. Two MVs are considered to be similar if the distance is below a threshold, ε_4. Since we know those 8×8 blocks have different motion or belong to another object, we should avoid assigning the same MV. Hence, the VMF chooses the most probable one among the candidate MVs that have passed the similarity check. In general, MV selection finds the object motion for a merged group with larger block size while motion refinement can be seen as local motion adjustment. Before updating $\mathbf{v}_{m,n}^*$ in MVF^2, we perform energy check on BPD of new obtained MV, $\mathbf{v}_{m,n}^*$, to ensure that its error energy is less than the original one, $\mathbf{v}_{m,n}$. Depending on how structure information is distributed on a 8×8 block, $\mathbf{v}_{m,n}^*$ may fail in the energy check. In such the case, we will leave it unchanged and try to correct it in the next iteration with an updated MVF or use MV smoothing process in [3] to reduce visual artifacts due to high BPD. That is, we set an iteration number to be 2 in motion refinement stage and after this stage, MV smoothing process is adopted as our last step for MV processing to obtain even finer MVF, MVF^3.

Generally, the frame interpolation scheme can be represented as:

$$f_t(i,j) = w_f f_{t-1}(i+\frac{1}{2}v_x, j+\frac{1}{2}v_y) + w_b f_{t+1}(i-\frac{1}{2}v_x, j-\frac{1}{2}v_y)$$

where f_{t-1} and f_{t+1} are two consecutive reconstructed frames, f_t is the interpolated frame, and $\mathbf{v} = (v_x, v_y)$ is the received MVF. w_f and w_b are the weights for the forward and backward predictions, respectively, which are often set to be 0.5. However, if a new object appears in the next frame and the previous frame only has part of the content, simply averaging bidirectional predictions will cause visual artifacts easily. Hence, for those MBs on the frame boundary, we propose using unidirectional interpolation based on the directions of their MVs. That is, we adaptively change the weights, w_f and w_b, which can be summarized as in Table 1.

Fig. 3. The interpolated results of frame 288 of WALK sequence using (a) original frame, (b) MV smoothing (PSNR: 19.76dB), (c) vector median filtering (PSNR: 19.67dB), and (d) the proposed method (PSNR: 21.33dB), respectively.

Fig. 4. The interpolated results of frame 56 of FORMULA 1 sequence using (a) original frame, (b) MV smoothing (PSNR: 29.46dB), (c) vector median filtering (PSNR: 29.80dB), and (d) the proposed method (PSNR: 31.56dB), respectively.

4. SIMULATIONS

In this section, we present experimental results to evaluate the performance of the proposed method. We compare our method with VMF, and MV smoothing as described in [3]. Four video sequences, FORMULA 1, WALK, BUS, and FAST FOOD, of CIF frame resolution are used and encoded using H.263 but even frames are skipped.

The averaged PSNR for these four video sequences are presented in Table 2, which shows our PSNR performance is consistently better than other methods. The visual comparisons are illustrated in Fig. 3 and Fig. 4. The result using the proposed method in Fig. 3(d) does not have blockiness artifacts as in Fig. 3(c). The blockiness artifacts can be removed by the method in [3] as shown in Fig. 3(b). However, ghost artifacts are generated due to impacts of incorrect MVs during the smoothing process. As one observes, Fig. 3(d) does not have the ghost effect since the proposed method corrects those unreliable MVs before smoothing. Moreover, we do not have broken structure around face and backpack such as Figs. 3(b) and (c) since we take edge information into consideration during the MV process. For FORMULA 1 sequence, fast motion and similar intensity of luminance between grass and pavement account for the failed interpolation on the white lines as shown in Figs. 4(a) and (b). However, these broken pieces do not appear in Fig. 4(d) and the proposed algorithm outperforms these two methods by removing most of artifacts. Also, our PSNR values are the highest in Fig. 3 and Fig. 4. More results can be found in http://videoprocessing.ucsd.edu/∼aihuang/.

Table 2. PSNR performance comparisons among two frame interpolation methods and the proposed method for four video sequences.

Sequences	Direct	VMF	Smooth	Proposed
WALK	22.88	22.95	22.99	**23.11**
FORMULA 1	28.09	28.24	27.85	**28.36**
BUS	22.92	23.03	22.64	**24.16**
FAST FOOD	25.67	25.84	25.69	**26.66**

5. CONCLUSIONS

We propose a novel algorithm for MCFI. Based on the received information, we first analyze the reliability levels of MVF and correct unreliable motion by finding major movement, which will be further refined with smaller block sizes. By doing this, we not only can accomplish the concept of object motion without complex motion estimation but we also can eliminate blocking artifact using smaller block size. Moreover, our method outperforms other conventional methods on both objective and subjective video quality.

6. REFERENCES

[1] S. Fujiwara and A. Taguchi, "Motion-compensated frame rate up-conversion based on block matching algorithm with multi-size blocks," *in Proc. Intl. Symp. on Intelligent Signal Processing and Communication Systems*, pp. 353–356, Dec. 2005.

[2] L. Alparone, M. Barni, F. Bartolini, and V. Cappellini, "Adaptively weighted vector-median filters for motion-fields smoothing," *in Proc. ICASSP'96*, vol. 4, pp. 2267–2270, May 1996.

[3] G. Dane and T. Q. Nguyen, "Smooth motion vector resampling for standard compatible video post-processing," *in Proc. Asilomar Conf. Signals, Systems and Computers*, 2004.

[4] A.-M. Huang and T. Nguyen, "Motion vector processing based on residual energy information for motion compensated frame interpolation," *in Proc. IEEE ICIP'06*, 2006.

[5] A.-M. Huang and T. Nguyen, "A novel motion compensated frame interpolation based on block-merging and residual energy," *in Proc. IEEE MMSP'06*, 2006.